IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ujiith Yogeswaran
<Date>

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- Data Collection Methodology Summary:
  - Data Collection
  - Data Wrangling
  - EDA with data visualization
  - EDA with SQL
  - Building an interactive map with Folium
  - Building a Dashboard with Plotly Dash
  - Predictive analysis (Classifications)
- Results Summary
  - Visual analysis
  - Interactive analytics
  - Predictive analysis

Introduction

Project summary and objectives:

- This project explores the commercial space industry, focusing on key players and their innovations. It aims to analyze the reuse potential of SpaceX’s Falcon 9 first stage using machine learning to provide insights for Space Y's strategic decisions.

Analyze the impact of commercial space companies:

- Investigate how companies like SpaceX, Virgin Galactic, Rocket Lab, and Blue Origin are transforming space travel by examining their business models, technological advancements, and contributions to making space travel more affordable and accessible.

Predict the reuse of SpaceX's Falcon 9 first stage:

- Use machine learning to predict the likelihood of the Falcon 9's first stage being reused, analyzing historical launch data and relevant variables to estimate future launch costs.

Provide insights for Space Y's competitive strategy:

- Gather and analyze data on SpaceX’s operations and costs to create dashboards. Use these insights to help Space Y develop a competitive edge by optimizing launch costs and improving its market position.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected by requesting the SpaceX API and cleaning the data to follow the correct format that is readable
- Perform data wrangling
  - The data was converted into a data frame and scatter plots and bar charts were made by writing Python code to analyze data and conducting exploratory data analysis
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Data was split into training data and test data to find the best Hyperparameter for SVM, Classification Trees, and Logistic Regression

Data Collection and Web Scraping

Data Collection:

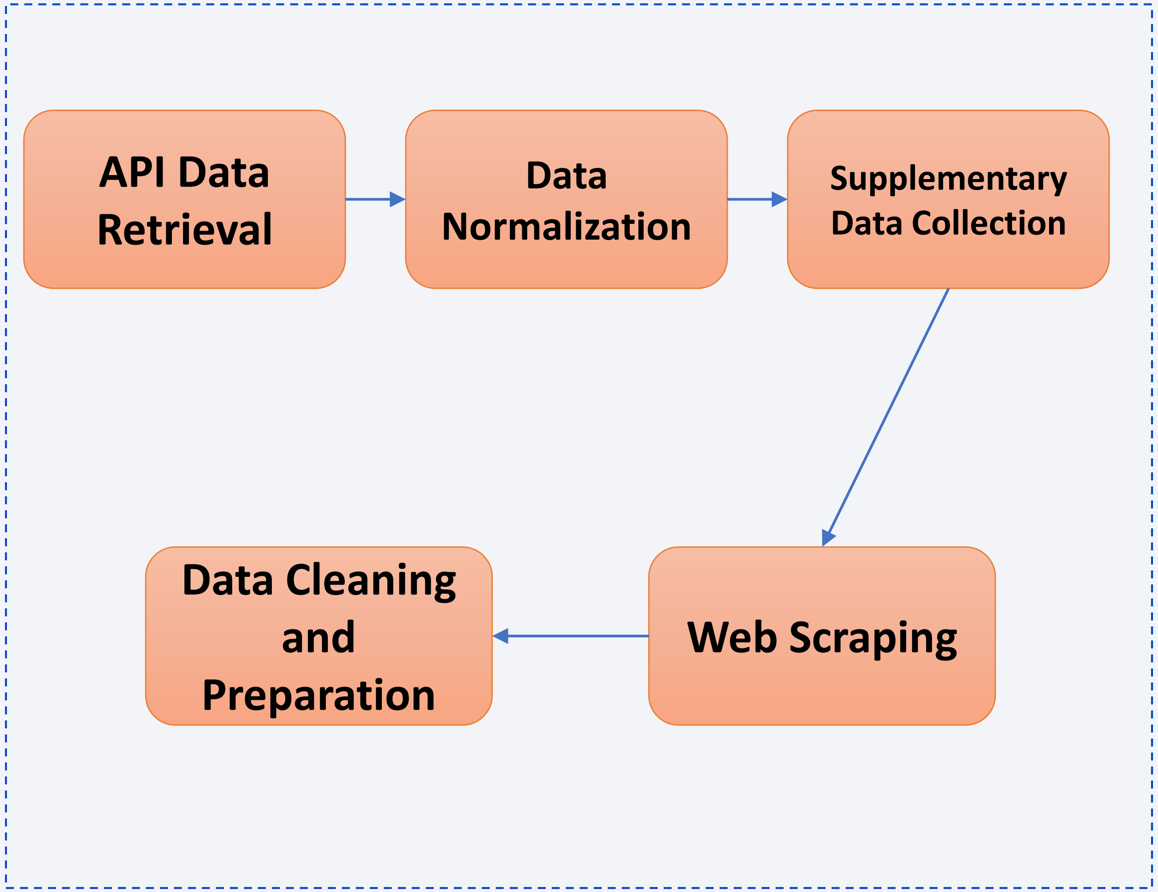
- Collected SpaceX launch data from the SpaceX REST API, targeting the api.spacexdata.com/v4/launches/past endpoint.
- Converted the JSON data into a flat table.
- Obtained additional details by querying other API endpoints for specific identifiers.
- Web scraped relevant Wikipedia pages using the BeautifulSoup package.

Data Cleaning:

- Filtered out Falcon 1 launches to focus on Falcon 9 data.
- Handled null values in the PayloadMass column by replacing them with the mean.
- Prepared LandingPad columns with null values for later one-hot encoding.

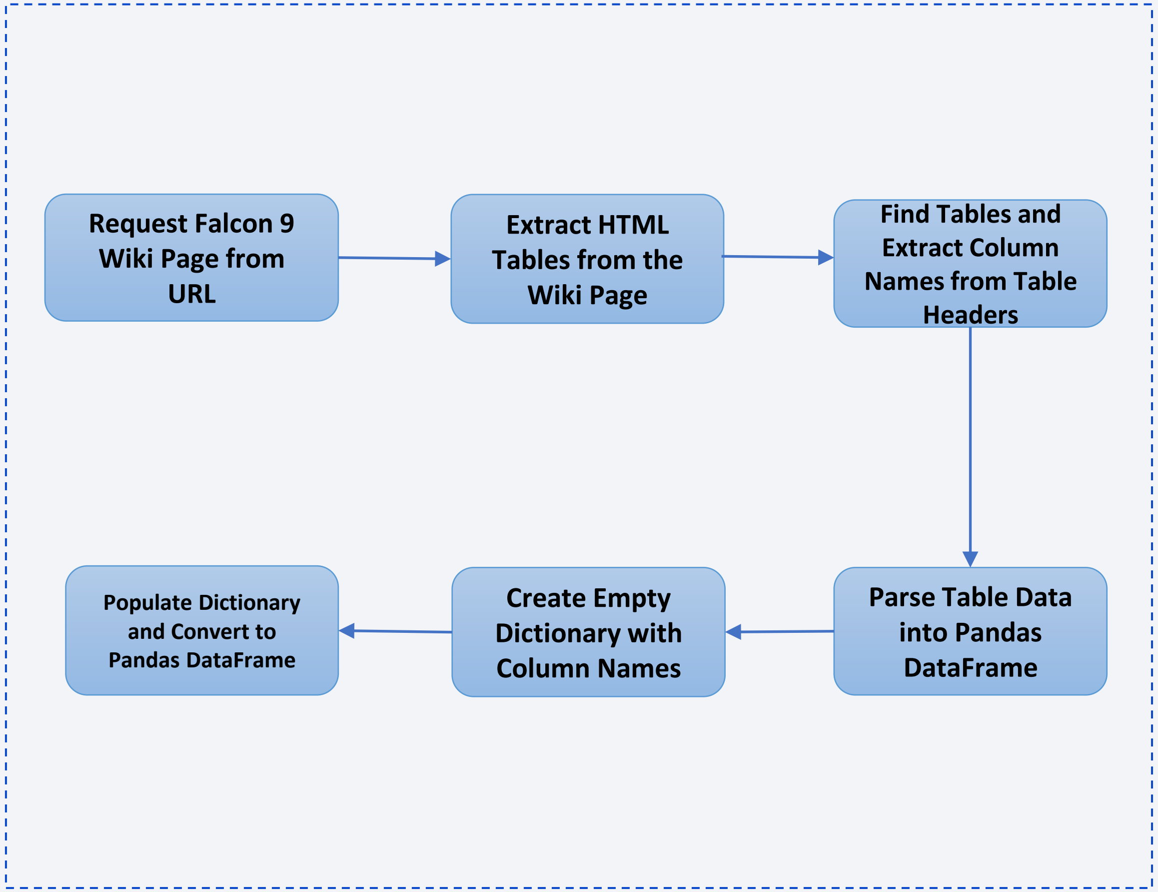
Data Collection – SpaceX API

- https://github.com/UYogesh3/Capstone-Project-Ujiith/blob/main/jupyter-labs-spacex-data-collection-api(1).ipynb



Data Collection - Scraping

- https://github.com/UYogesh3/Capstone-Project-Ujiith/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Reviewed Key Attributes:
  - Analyzed various attributes of the SpaceX launch data, including Flight Number, Date, Booster Version, Payload Mass, Orbit, Launch Site, and Outcome.
  - Examined Specific Columns:
  - "LaunchSite": Identified and reviewed the unique launch sites to understand the geographical distribution of launches.
  - "Orbit": Investigated different orbital destinations such as Low Earth Orbit (LEO) and Geostationary Transfer Orbit (GTO) to categorize payload missions.
- Analyzed the Outcome Column:
  - Focused on the "Outcome" column, which indicates the landing status of the Falcon 9 first stage.
- Converted Outcome to Binary Classification:
  - Transformed the "Outcome" column into binary classes
  - 0 for unsuccessful landings
  - 1 for successful landings
  - This binary classification was used to standardize the target variable for analysis.
- Prepared Data for Prediction:
  - Used the binary classification (YY) to set up the target variable for predicting launch outcomes in the analysis

.https://github.com/UYogesh3/Capstone-Project-Ujiith/blob/main/labs-jupyter-spacex-Data%20wrangling(1).ipynb

EDA with Data Visualization

- Flight Number vs. Payload:
  - Visualized the relationship between flight number and payload to understand how payload mass affects launch outcomes over successive launches.
- Flight Number vs. Launch Site:
  - Examined the relationship between flight number and launch site to analyze how launch frequency varies across different locations.
- Payload vs. Launch Site:
  - Illustrated the relationship between payload mass and launch site to identify patterns in payload distribution at different sites
- Bar Chart: Success Rate by Orbit Type:
  - Displayed the success rate for each orbit type to compare performance and effectiveness across different orbital missions.
- Flight Number vs. Orbit Type:
  - Analyzed the relationship between flight number and orbit type to assess how launch frequency impacts different types of orbits.
- Payload vs. Orbit Type:
  - Investigated the relationship between payload mass and orbit type to understand how payload mass influences success rates in various orbits.
- Line Chart: Yearly Launch Success Trend:
  - Plotted the yearly trend of launch success to track changes and improvements in launch outcomes over time

https://github.com/UYogesh3/Capstone-Project-Ujiith/blob/main/edadataviz.ipynb

EDA with SQL

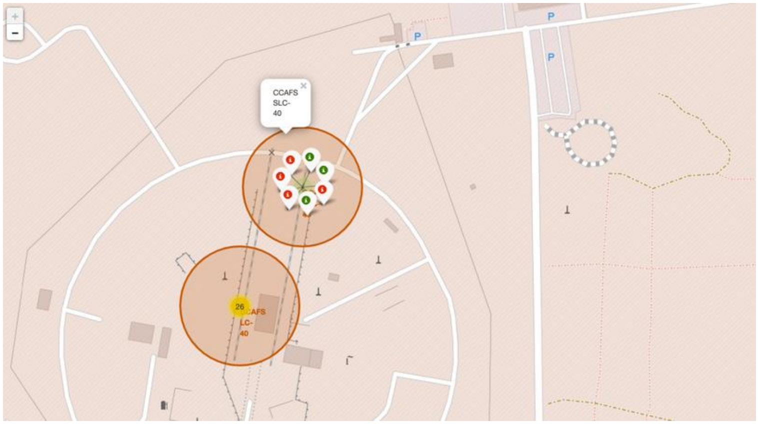
- Task 1: Display the names of the unique launch sites in the space mission.
- Task 2: Display 5 records where launch sites begin with the string 'CCA'.
- Task 3: Display the total payload mass carried by boosters launched by NASA (CRS).
- Task 4: Display the average payload mass carried by booster version F9 v1.1.
- Task 5: List the date when the first successful landing outcome on ground pad was achieved (using the min function).
- Task 6: List the names of the boosters which have success on a drone ship and have payload mass between 4000 and 6000.
- Task 7: List the total number of successful and failed mission outcomes.
- Task 8: List the names of the booster versions which have carried the maximum payload mass (using a subquery).
- Task 9: List the records displaying month names, failure landing outcomes on a drone ship, booster versions, and launch sites for the months in the year 2015 (using substr function for month and year extraction).
- Task 10: Rank the count of landing outcomes (e.g., Failure (drone ship) or Success (ground pad)) between 2010-06-04 and 2017-03-20 in descending order.

EDA with SQL

- https://github.com/UYogesh3/Capstone-Project-Ujiith/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

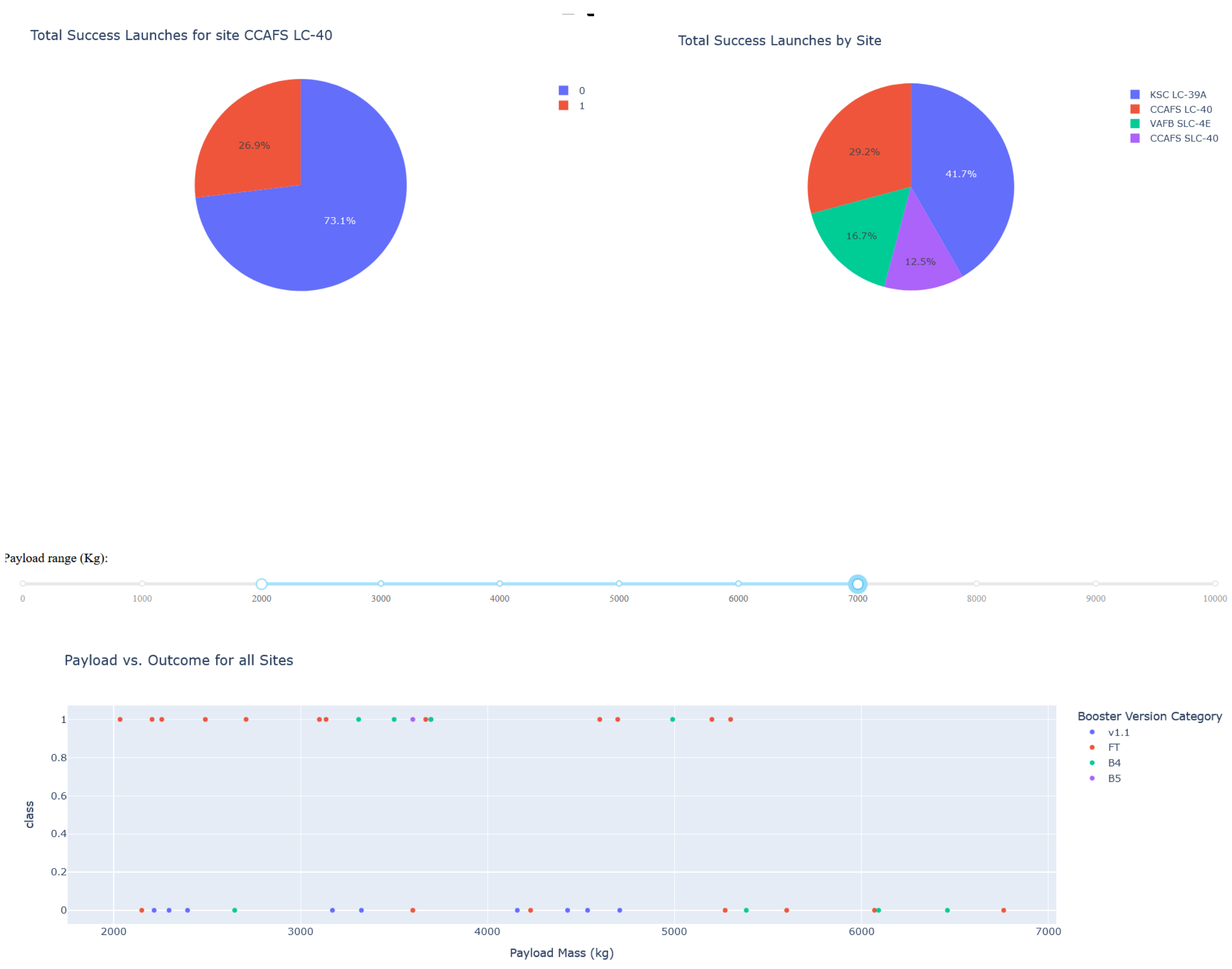
Build an Interactive Map with Folium

- https://github.com/UYogesh3/Capstone-Project-Ujiith/blob/main/lab_jupyter_launch_site_location.ipynb



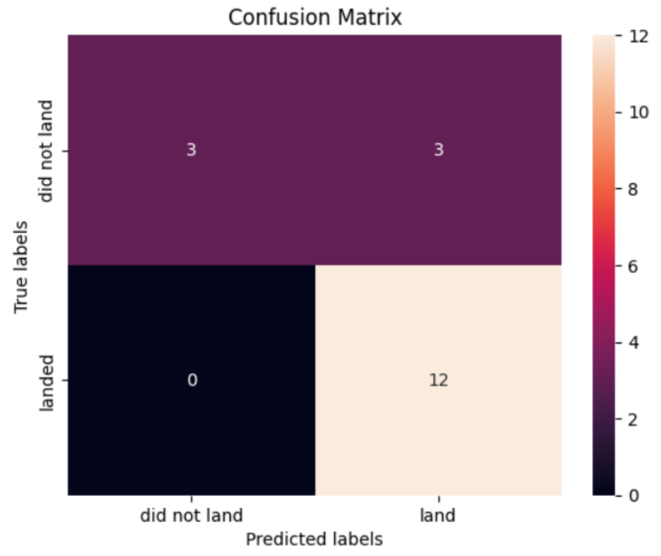
Build a Dashboard with Plotly Dash

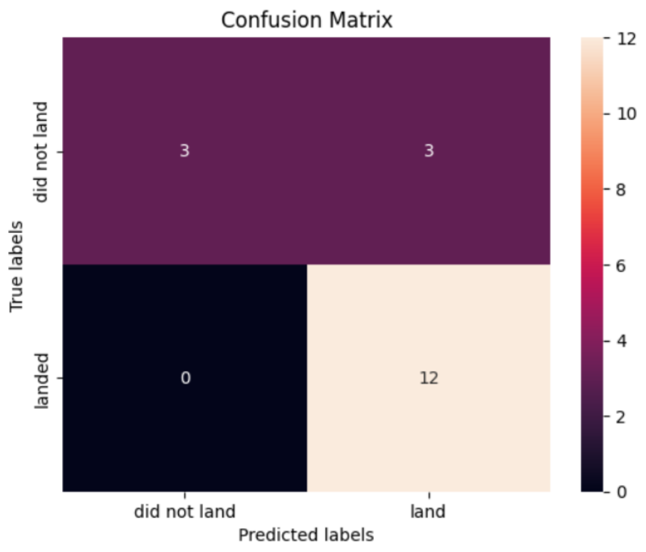
- https://github.com/UYogesh3/Capstone-Project-Ujiith/blob/main/spacex_dash_app.py

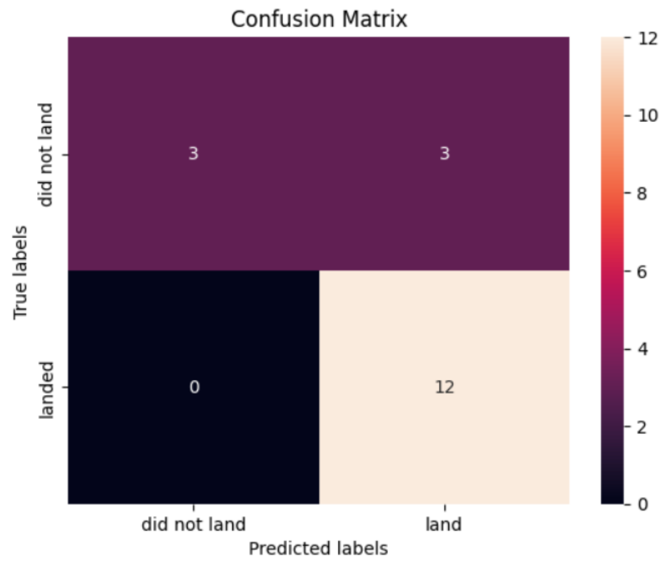


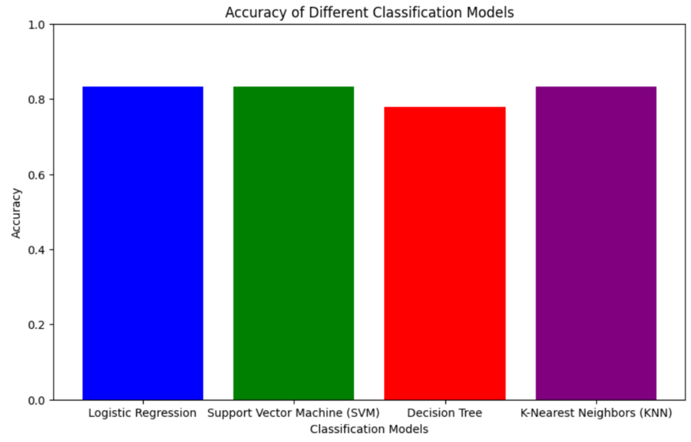
Predictive Analysis (Classification)

- https://github.com/UYogesh3/Capstone-Project-Ujiith/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb







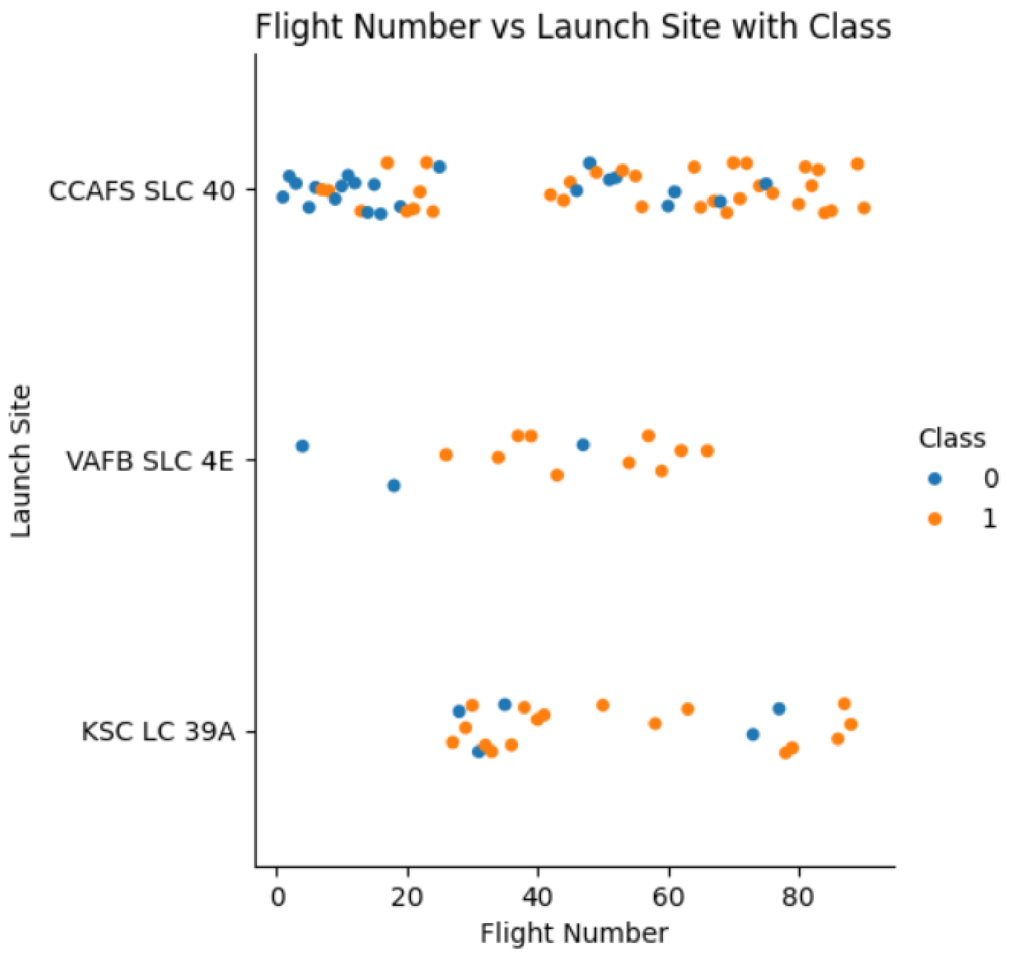


Results

- Lower Payload mass have a higher chance of success than missions with higher Payload mass
- KGS LC 39 has the most successful missions
- The orbits with the most success are HEO, GEO, ESL, and SSO
- For classification accuracy the models with the highest accuracy are Logistic regression, SVM, and KNN

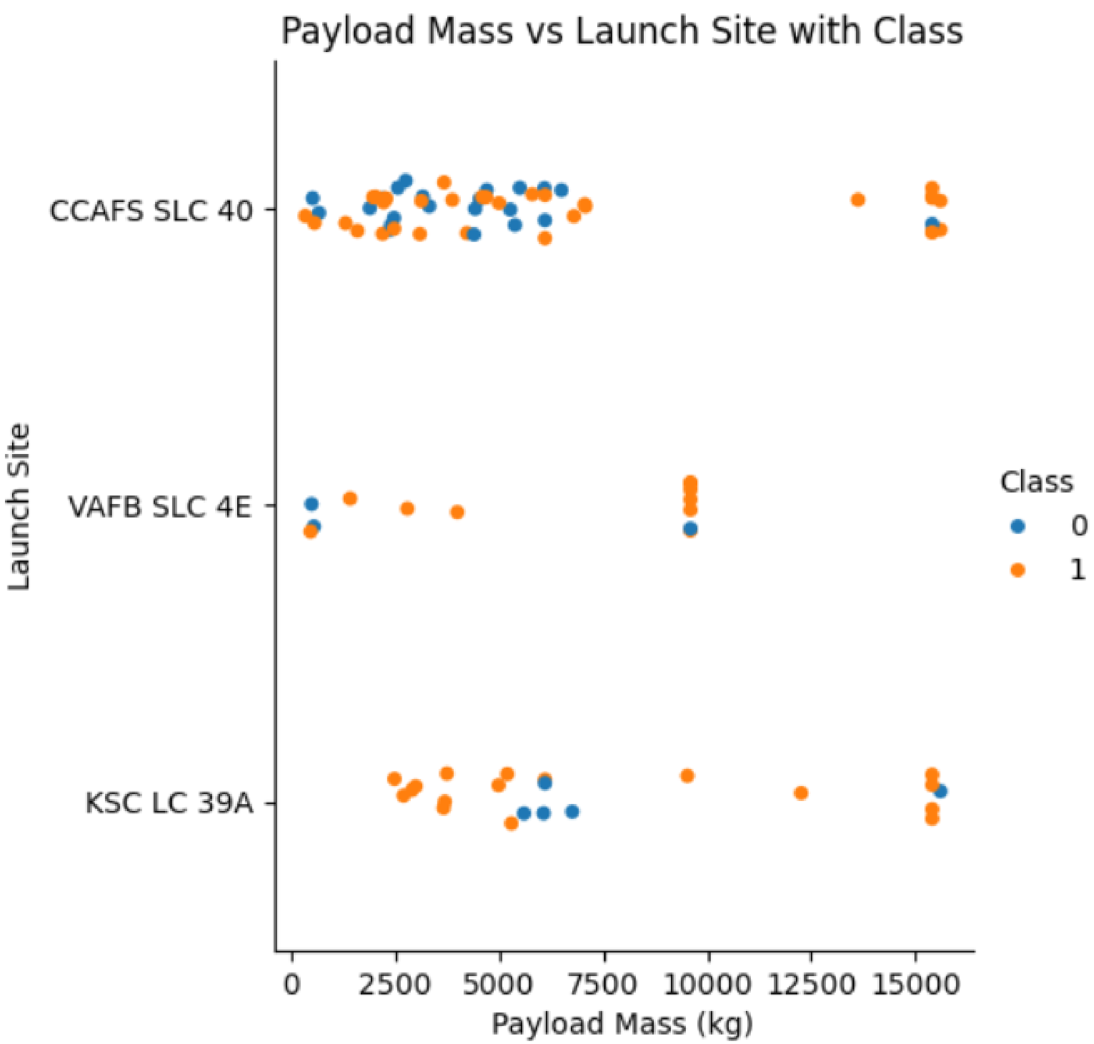
Section 2

Insights drawn from EDA



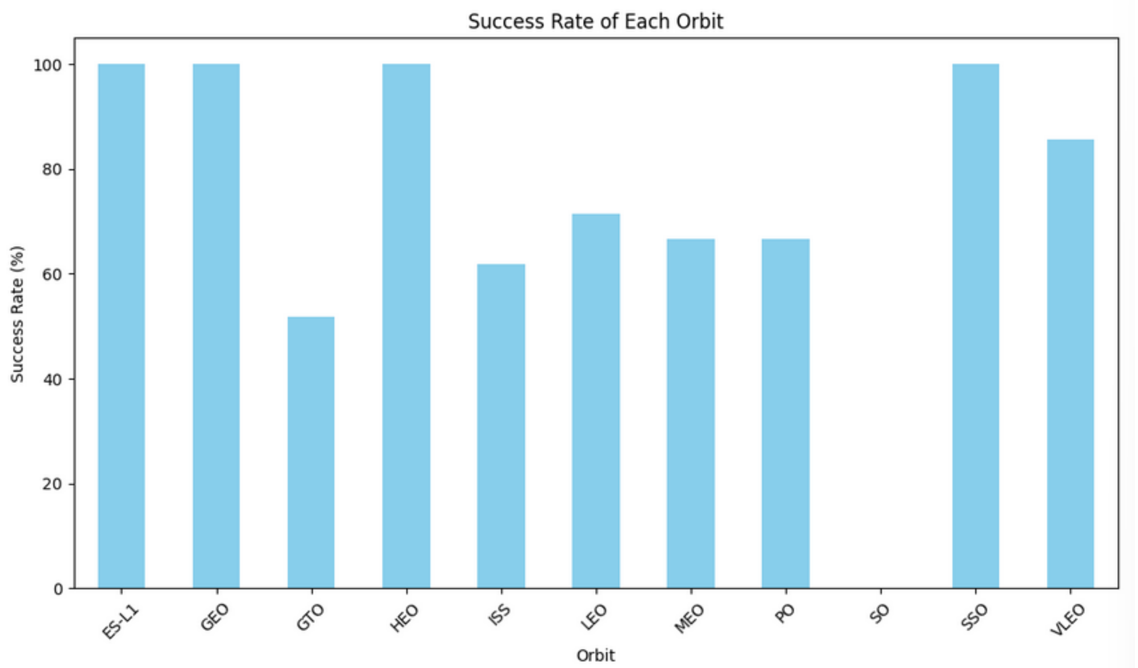
Flight Number vs. Launch Site

- Launch site CCAFS SLC 40 had the most success in relation to flight numbers while VAFB SLC 4E had the least success and failures overall. Launch site KSC LC 39A had more success in higher flight numbers



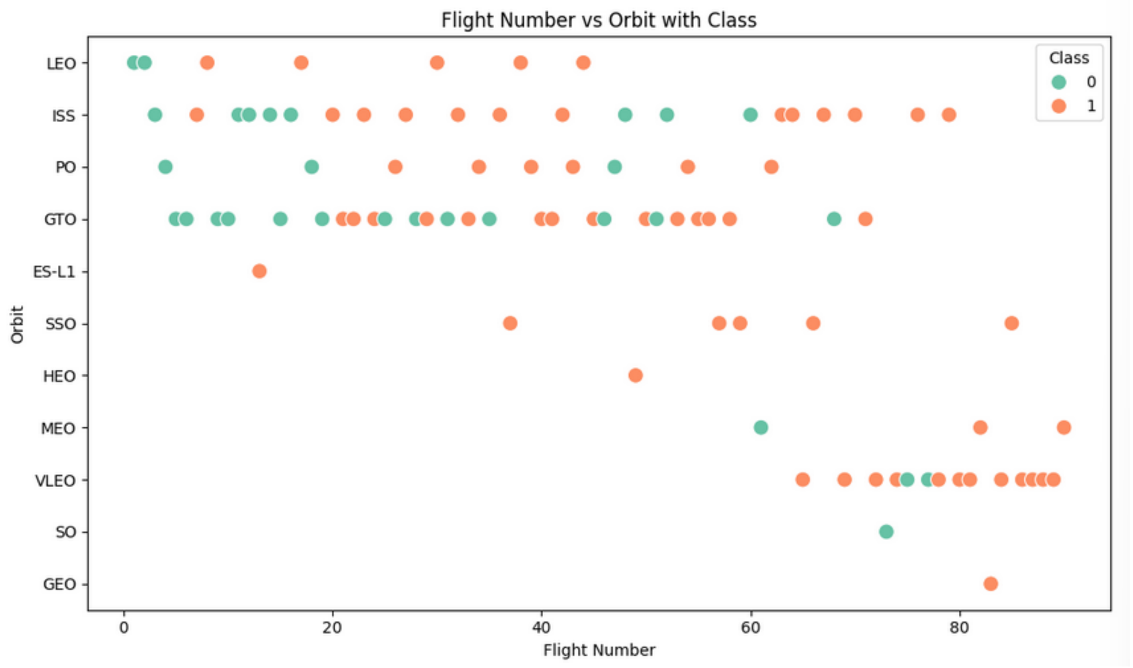
Payload vs. Launch Site

- for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).



Success Rate vs. Orbit Type

- The success rate was highest on orbits ES-L1, GEO, HEO, AND SSO while the lowest rate of success was orbit SO

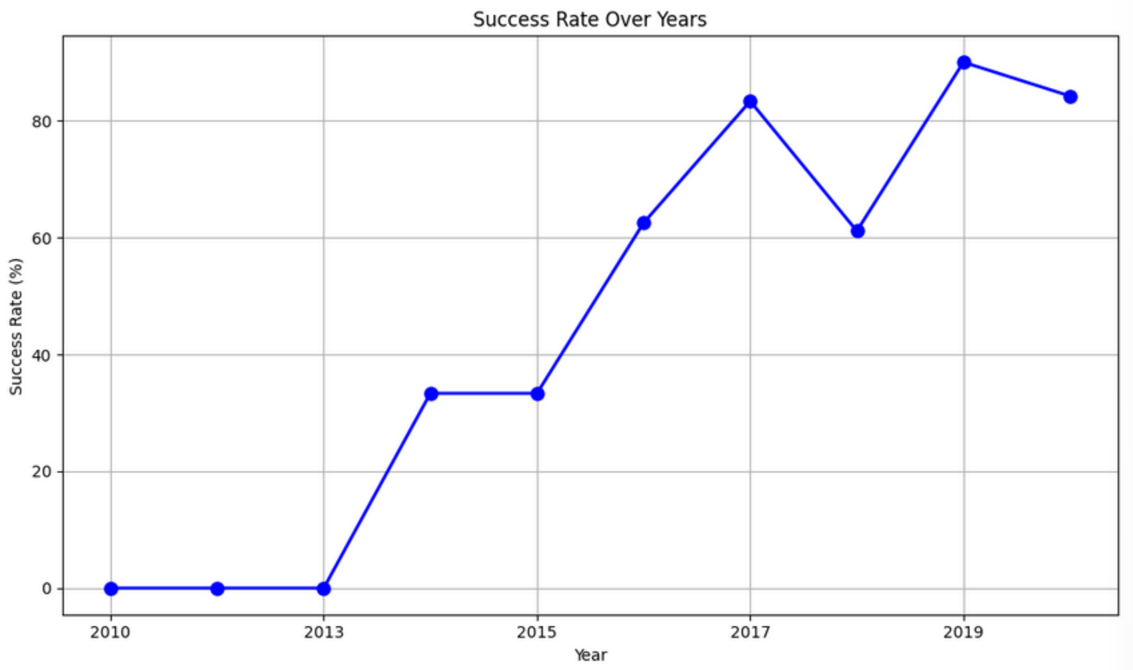


Flight Number vs. Orbit Type

- Orbit ISS had the most consistent successes with increasing flight numbers, orbits ES-L1, SSO, HEO, and GEO all had success regardless of having minimal attempts. Orbit VLEO had the highest success in higher flight numbers

Payload vs. Orbit Type

- Most of faliures fall between 2000 kg and 8000kg mainly between orbits ISS and GTO. Orbits ES-L1, SSO, HEO, and GEO all had successes between 0 and 6000kg



Launch Success Yearly Trend

- The overall trend of the graph shows the success rate has increased over the year with exception of 2018. The highest rate of success was between 2015 and 2017

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

All Launch Site Names

- These are the individual unique launch sites that were found with the given data
- %sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;

Launch Site Names Begin with 'CCA'

- By filtering out launch sites started with ‘CCA’ we were able to find 5 records where noticeably 2 out of the 5 had successful mission outcomes with no payload mass yet failed in the landing
- %sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

total_payload_mass
45596

- %sql SELECT SUM("Payload_Mass__kg_") AS Total_Payload_Mass FROM SPACEXTABLE WHERE "Mission_Outcome" LIKE '%CRS%';

Average Payload Mass by F9 v1.1

Average_Payload_Mass
2928.4

- Calculated the average payload mass carried by booster version F9 v1.1
- %sql SELECT AVG(PAYLOAD_MASS__KG_) AS Average_Payload_Mass FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1';

First_Successful_Landing_Date
2015-12-22

First Successful Ground Landing Date

- the first successful landing outcome on ground pad was December 22, 2015
- %sql SELECT MIN(Date) AS First_Successful_Landing_Date FROM SPACEXTABLE WHERE Landing_Outcome LIKE '%Success (ground pad)%';

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Successful Drone Ship Landing with Payload between 4000 and 6000

- These are the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
- %sql SELECT DISTINCT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome LIKE '%Success (drone ship)%' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;

Total Number of Successful and Failure Mission Outcomes

Mission_Outcome	Total_Count
Success	98

- Out of the total number of attempts only 98 attempts were successful mission outcomes
- %sql SELECT Mission_Outcome, COUNT(*) AS Total_Count FROM SPACEXTABLE WHERE Mission_Outcome IN ('Success', 'Failure') GROUP BY Mission_Outcome;

Boosters Carried Maximum Payload

- These are all the F9 B5 boosters that carried the most payload in KG
- %sql SELECT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);

Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

Month_Name	Landing_Outcome	Booster_Version	Launch_Site
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

2015 Launch Records

- These are the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- %sql SELECT CASE WHEN substr(Date, 6, 2) = '01' THEN 'January' WHEN substr(Date, 6, 2) = '02' THEN 'February' WHEN substr(Date, 6, 2) = '03' THEN 'March' WHEN substr(Date, 6, 2) = '04' THEN 'April' WHEN substr(Date, 6, 2) = '05' THEN 'May' WHEN substr(Date, 6, 2) = '06' THEN 'June' WHEN substr(Date, 6, 2) = '07' THEN 'July' WHEN substr(Date, 6, 2) = '08' THEN 'August' WHEN substr(Date, 6, 2) = '09' THEN 'September' WHEN substr(Date, 6, 2) = '10' THEN 'October' WHEN substr(Date, 6, 2) = '11' THEN 'November' WHEN substr(Date, 6, 2) = '12' THEN 'December' END AS Month_Name, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTABLE WHERE substr(Date, 0, 5) = '2015' AND Landing_Outcome LIKE '%Failure (drone ship)%';

Landing_Outcome	Outcome_Count
Failure (drone ship)	5
Success (ground pad)	3

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

- %SELECT Landing_Outcome, COUNT(*) AS Outcome_Count FROM SPACEXTABLE WHERE Date_Time_UTC BETWEEN '2010-06-04' AND '2017-03-20' AND Landing_Outcome IN ('Failure (drone ship)', 'Success (ground pad)’) GROUP BY Landing_Outcome ORDER BY Outcome_Count DESC;

Section 3

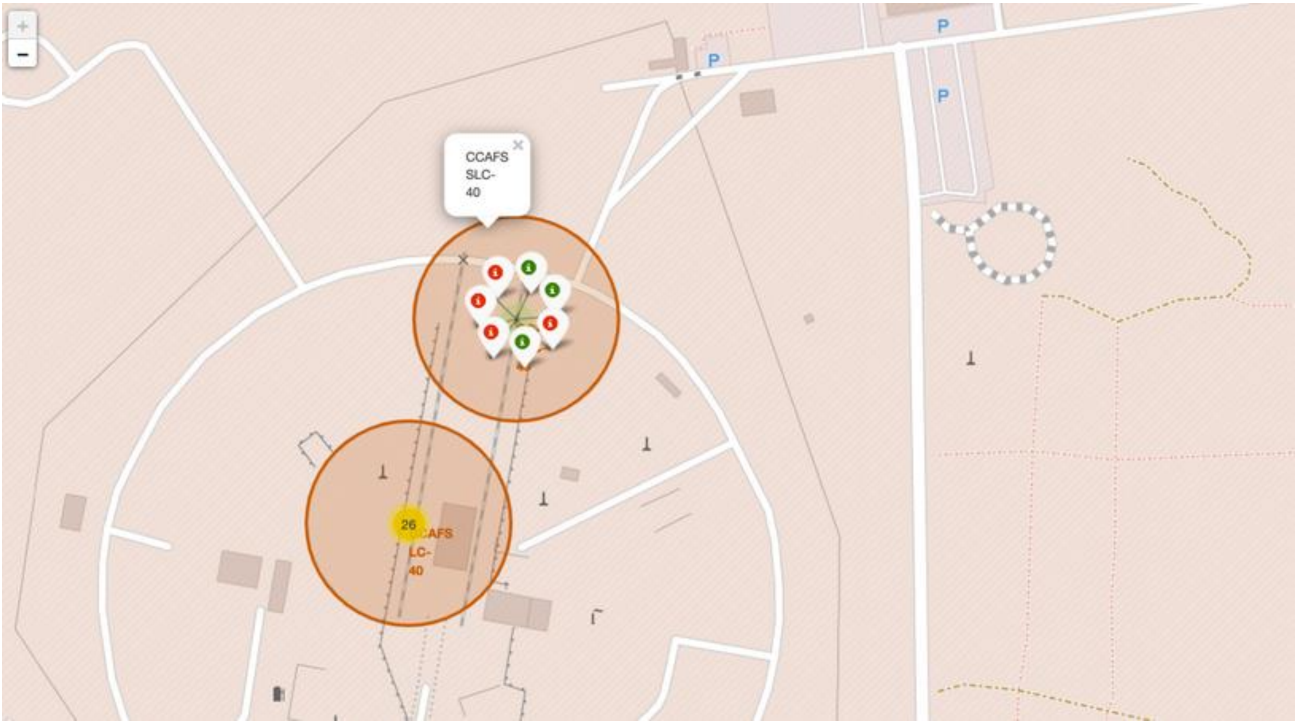
Launch Sites Proximities Analysis

Folium Map

- Following are the launch site locations

Folium Map

- Clusters of launch sites red meaning they failed while green meaning the launches were a success



Folium Map

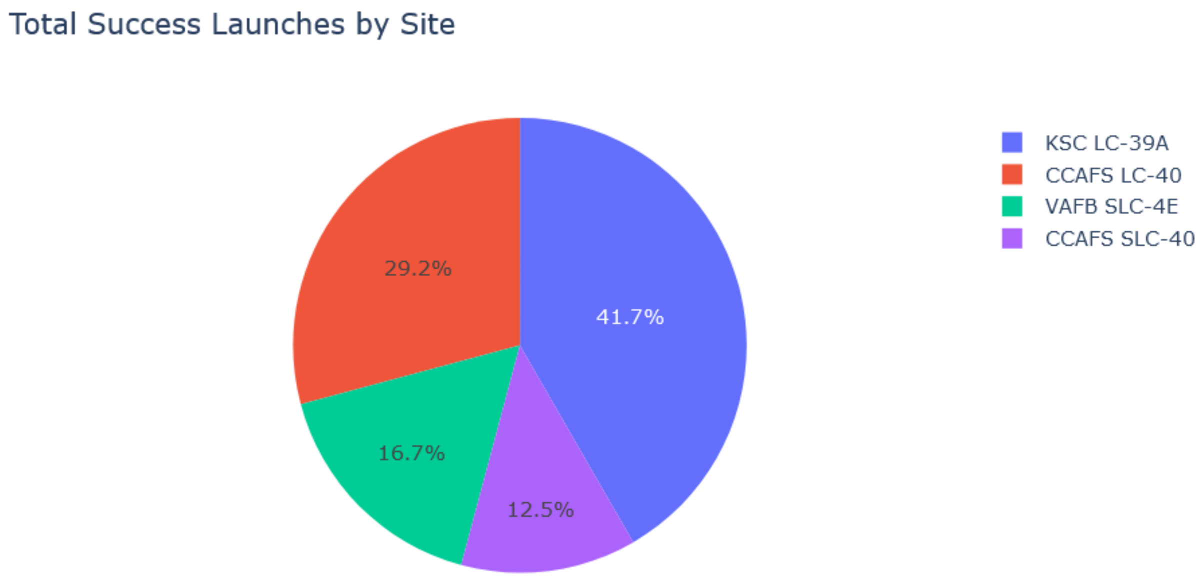
- All launch sites are within safe distance of cites and located near railways, highways and coastlines

Section 4

Build a Dashboard with Plotly Dash

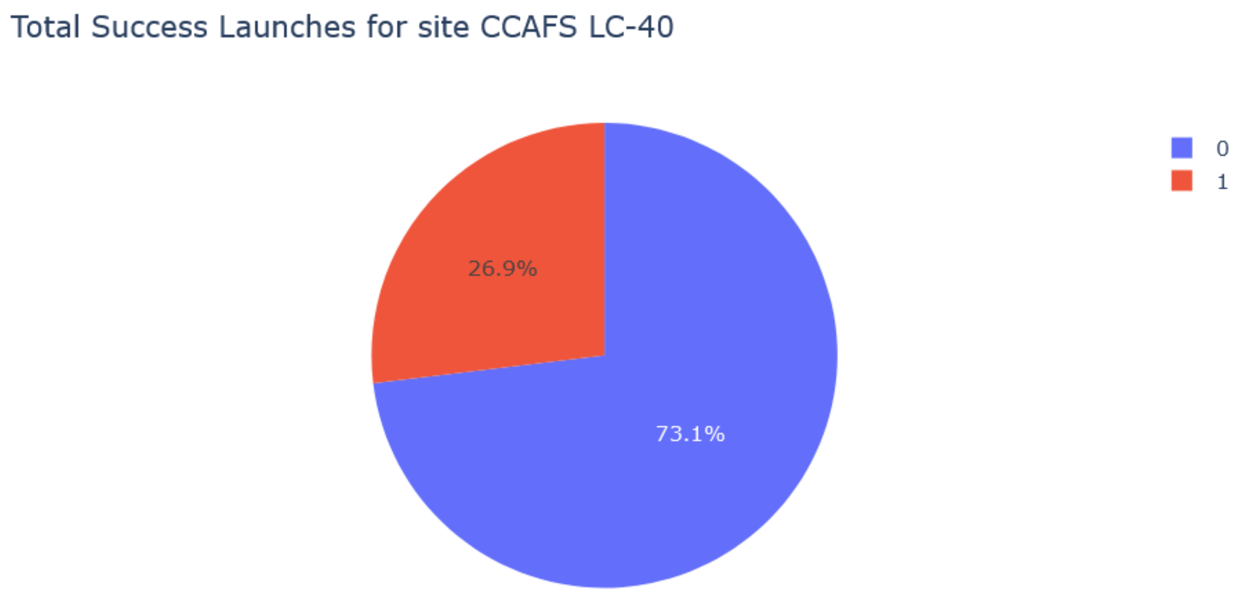
Dashboard

- Sites KSC LC 39A and CCAFS LC-40 has the most successes overall with site CCAFS SLC-40 having the least success



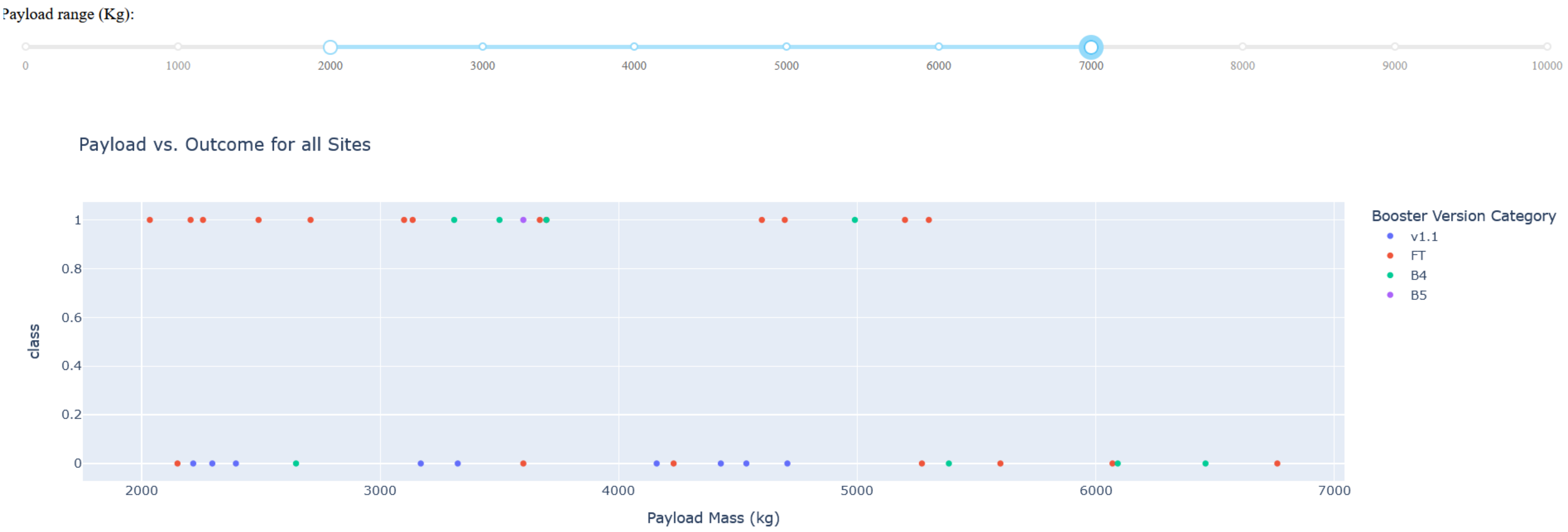
Dashboard

- For on of the more successful sites only 26.9% of the attempts failed



Dashboard

- Low payload mass favors higher success rates than higher payloads

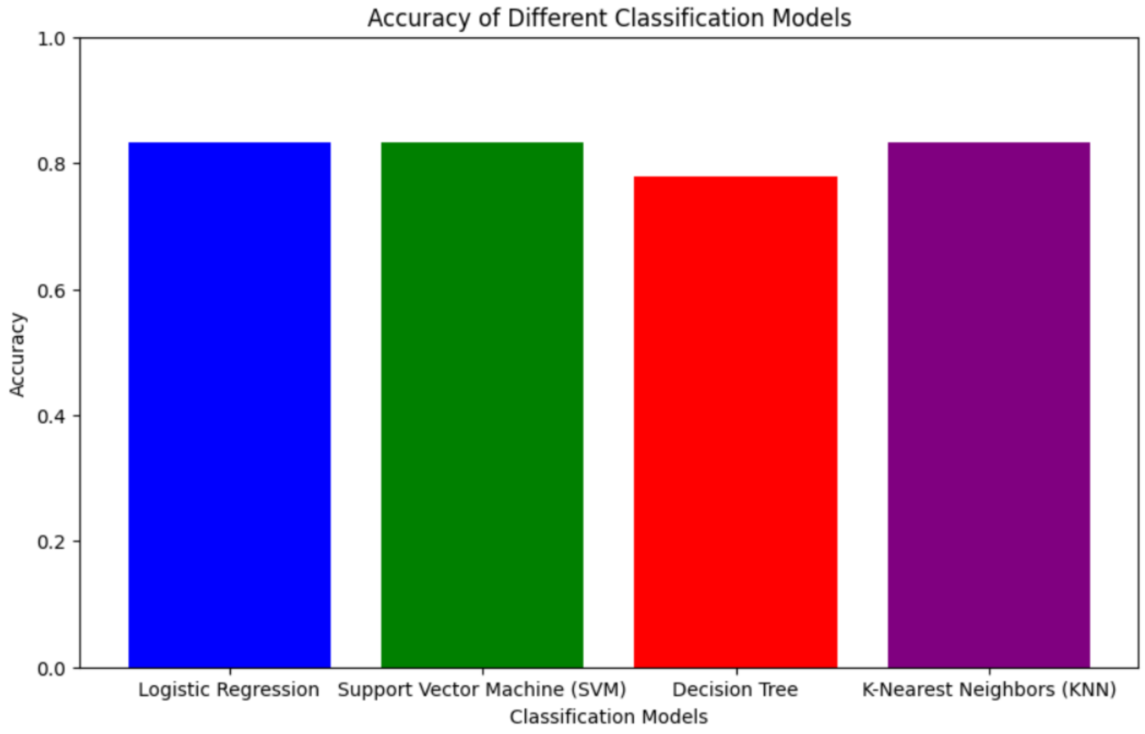


Section 5

Predictive Analysis (Classification)

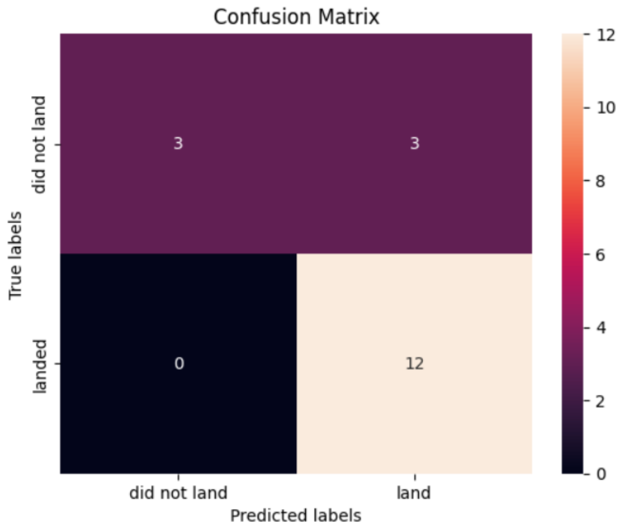
Classification Accuracy

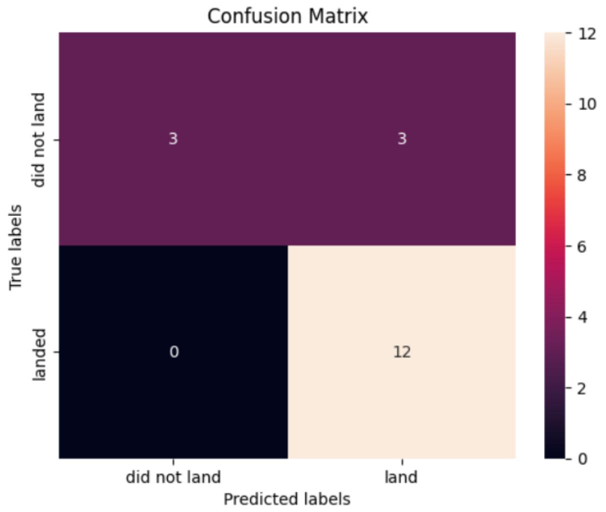
- The best classification models were between Logistic regression, SVM, and KNN

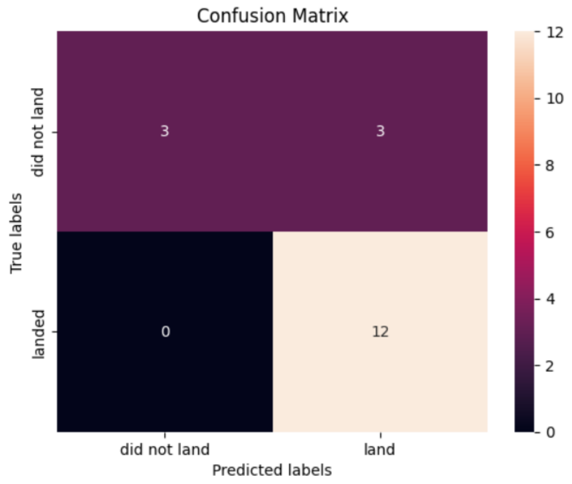


Confusion Matrix

The 3 classification models that has the most success







Conclusions

- Lower Payload mass have a higher chance of success than missions with higher Payload mass
- KGS LC 39 has the most successful missions
- The orbits with the most success are HEO, GEO, ESL, and SSO
- For classification accuracy the models with the highest accuracy are Logistic regression, SVM, and KNN

Appendix

- Main Capstone Project Github:
- https://github.com/UYogesh3/Capstone-Project-Ujiith/tree/main

Thank you!